DWD Use Cases

J. S. Marron, Xuxin Liu, Everett Zhou

Use case 1: How DWD is currently used

Researcher X purchases and installs Matlab on his desktop. He downloads the current Matlab implementation of DWD from the web site https://genome.unc.edu/pubsup/dwd/, and installs it, as instructed in the file ReadMe.txt. He starts Matlab, and reads his data into a Matlab matrix. While Matlab can accept either tab-delimited text files, or Excel spreadsheets, care is needed because handling mixed text and numeric data in Matlab can be a slippery task. A single call to the Matlab program BatchAdjustSM.m does the DWD re-centering step, and returns a data matrix and some diagnostic plots. Often it is desirable to standardize the data columns, which is easily done in Matlab. The data are then written to a file.

Relevant issues for Architecture:

1. Data to Matlab matrix conversion is tricky and case-wise.
2. Local installation of Matlab, SDPT3 and in our package: In a federated system, all local installation should be avoided if possible.
3. Tracking data manipulations: all data manipulations and workflows should be tracked automatically.
4. Need convenient output format.

Use case 2: How we plan to use DWD

Researcher X launches the DWD client. He loads in the two data sets to be combined, in the form of two micro-array CDEs. The DWD client generates the adjusted data, and graphics that indicates how successful the DWD process has been at normalization. If the graphic suggests that a further column standardization step will be useful, the client allows this step as well, and studies the quality of the result with a further graphical display. The final combined data set is output as a new combined micro-array CDE.

Relevant issues for Architecture:

1. Appropriate, commonly agreed upon, micro-array CDE needs to be defined.
2. The output CDE should allow a new index (for the arrays), that indicates which source it came from.
3. The output CDE should allow annotation indicating the original data sets, that DWD was performed, and whether or not column standardization was done.

Use case 3: How we hope to use DWD

Researcher X wants to combine two different microarray data sets. The data are stored as caBIG CDEs, in one of the agreed upon formats. He goes to the caBIG webpage, and

finds the DWD page. He fills out the fields indicating the CDEs to be combined (or perhaps pulls these from a menu), and gives a name for the new CDE with the combined data, and the location on the grid for storage of the resulting new CDE. The web page produces some diagnostic plots, and asks if the result looks acceptable. If so, Researcher X pushes a button, which causes the new CDE to be created. If the results are unacceptable, Researcher X chooses from a menu of options for enhanced adjustment. After an iterative process, with each step being facilitated by appropriate visualization, the results are accepted, and the output CDE is generated.

Relevant issues for Architecture:

1. Seamless connection of related systems, such as data storage grid, web server.
2. Ability to read CDEs from any grid location.
3. Permission schema for allowing the writing of the new combined CDEs in appropriate locations.
4. Ability to save diagnostic graphics.